Organising Documentation in Knowledge Evolution and Communication

Cristina DE CASTRO, PAOLO TOPPAN
IEIIT-CNR, Italian National Research Council
V.le Risorgimento 2, 40136, Bologna, Italy

ABSTRACT

The knowledge of a subject evolves in time due to many factors, such as better understanding, study of additional issues within the same subject, study of related work from other themes, etc. This can be achieved by individual work, direct cooperation with other people and, in general, knowledge sharing. In this context, and in the broader context of knowledge communication, the appropriate organisation of documentation plays a fundamental role, but is often very difficult to achieve.
A layered architecture is here proposed for the development of a structured repository of documentation, here called *knowledge-bibliography KB*. The process of knowledge acquisition, evolution and communication is firstly considered, then the distributed nature of nowadays knowledge and the ways it is shared and transferred are taken into account. On the basis of the above considerations, a possible clustering of documentation collected by many people is defined. An LDAP-based architecture for the implementation of this structure is also discussed.

Keywords: Knowledge Evolution, Layered Architectures, Acquisition and Documentation, Sharing and Transferring, Knowledge-Bibliography, LDAP Architectures

1. INTRODUCTION

Most nowadays activities are based on the synergy of many different personal experiences and backgrounds, hence the process of communication has become an essential cornerstone of the advancement of knowledge. A wide experience in a certain subject involves many aspects, such as personal theoretical and practical skills, insights arising from different sources and so on. This material can be documented and formalised in many heterogeneous ways and memorised on different kinds of media or, on the contrary, it can be neither documented nor formalised at all. As a consequence, knowledge transfer [1, 2, 3, 4, 5] is a very complex process, in which the Internet and real-time communications play two antithetical roles: on the one hand, they make information highly and quickly available; on the other hand, they increase the need to properly organise this unlimited material and make it available in a structured, simple and meaningful layout. Many efforts have been made on this subject, some of whom can be found in [6, 7, 8, 9].

In this paper a layered architecture is discussed for the study and development of a user-friendly multimedia *knowledge-bibliography (KB)* on a specific subject and its related ones. This model is based on some basic considerations and schemes for representing the process of knowledge acquisition, evolution and communication by a single person or by many people together. An LDAP scheme is proposed for developing the resulting structure.

2. THE PROPOSED MODEL

The main requirement of the knowledge-bibliography is to provide a huge and structured variety of documentation on a specific subject and links to related issues. Furthermore, some main contacts deriving from direct communication or other types of learning and information sharing should be represented, in order to allow people find the experts in these themes. Such structure should also be easily shared and extended and, possibly, integrated in a broader architecture containing a wide range of different subjects.
This kind of structure can arise from the joint work of many people and, in all probability, from the integration of the documentation that each of them has been collecting over time. It must be noticed that the way a person organises and judges the relevance of some material is very subjective. As far as the organisation is concerned, a sort of layout could be imposed, as it happens in papers and books. On the contrary, as far as relevance is concerned, it is almost impossible (and maybe absolutely unfair) to leave aside personal opinions, even if many criteria have been stated (see for instance [10, 11, 12, 13]).

Far from expecting to solve this very difficult problem, this paper presents some simple schemes that lead from the process of learning to the process of communicating, in order to give some guidelines for creating a well organised documentation. This documentation arises from many different sources and persons and is translated into a common format.

A Layered Architecture for the Knowledge-Bibliography

Many representations of knowledge by means of graphs have been developed, among whom [14, 15, 16, 17, 18]. The perspective of the proposed architecture is communication and cooperation of many people about a certain subject, each person owning a personal knowledge. Adopting a layered graph representation, the top layer refers to subjects, the middle layer refers to people and the bottom layer represents the knowledge of each person on a specific subject. Each layer can be composed of many items, which can also be inter-connected.

This type of representation is known as *layered knowledge architecture (LK architecture)* and it is used in many application fields, ranging from path planning in structured environments to software engineering, the semantic web and many others [19, 20, 21, 22, 23]. Let us describe the LK model by means of an example, in order to adapt it to our purpose afterwards. Suppose you must describe an office in a palace with three floors, each floor containing some rooms and each room containing some office furniture, such as chairs, tables, lamps, PCs, etc. In the LK formalism, the office is represented by a multi-level graph. At the bottom level, named *symbolic layer*, the nodes (*landmarks*) are the pieces of furniture and the edges (*routes*) are the physical paths among them. Higher level layers can be defined by *aggregation*, *generalization* or *classification*. In the example, the first layer can be obtained by

aggregation, clustering the furniture of the same room. The second knowledge layer (*1-clustered layer*) is obtained by aggregation into rooms. The 1-clustered layer is thus a graph whose nodes are the rooms of the palace and whose edges are the paths connecting the rooms. In the same way, the *2-clustered layer* is obtained by aggregating the rooms of the same floor. In this way a graph is obtained whose nodes are the floors of the palace and whose edges are the stairs. In the *3-clustered layer* all the floors are aggregated in a single node representing the whole office.

As far as the knowledge-bibliography is concerned, the LK architecture is here slightly generalised and works as follows (Fig. 1): the symbolic layer contains the documentation items k_{ij}, which can be clustered together if they belong to the knowledge scope of a certain person and thus define the 1-clustered layer. Each item can, of course, belong to many different clusters. The people that share a common interest in the specific subject define the 2-clustered layer. Many levels of detail can be defined by creating further layers, clustering by sub-areas and so on.

In order to underline each person's contribute in a straightforward way, this representation must be initially simplified by hiding the links among the nodes belonging to each level. In particular, the links among the nodes of the symbolic layers that belong to different clusters are hidden and so are the links among people. In this way, some preferential nodes and paths are imposed to the initial graph (Fig.2), leading thus to a tree structure. A similar approach can be found in many pruning techniques and in [24].

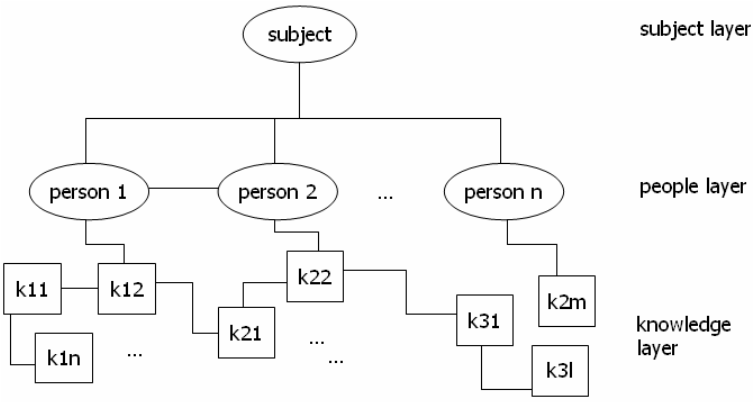

Fig. 1: basic 3-layer clustered architecture for the knowledge-bibliography

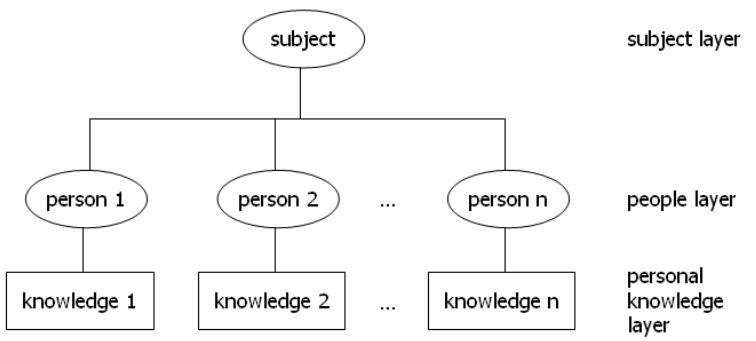

Fig. 2: the knowledge-bibliography structure from the viewpoint of personal knowledge

In order to define this structure in detail, the following steps are taken into account:

(a) First, the viewpoint of a single person on a single subject is considered

(b) This person collects his/her documentation on this subject and consequently build a *partial knowledge-bibliography (PKB)*, which arises from his/her cluster in the personal knowledge layer

(c) Due to the process of learning and knowledge communication among people involved in the same area, the personal knowledge evolves, and so does each partial knowledge-bibliography.

Further steps (maybe the most difficult ones) derive from the following considerations:

(d) Steps (a), (b), (c) lead to a network of partial knowledge-bibliography units

(e) Efforts ought to be made through cooperation and knowledge communication in order to make a knowledge-bibliography (KB) on the subject or, best, on a knowledge area

(f) Step (e) could be used to create a higher level network of different related knowledge areas.

Steps (a), (b), (c) will be discussed in the following subsections.

How documentation can evolve through knowledge evolution and communication

Let us consider a single person P_i, a single subject S_j and P_i's partial knowledge-bibliography PKB_i. The process of knowledge acquisition, evolution and communication starts at the *knowledge state* st_1 and can be roughly depicted as shown in Fig. 3: P_i acquires information about S_j from some initial sources, then the process of learning starts. P_i can produce both digital documentation and non-documented knowledge (*partial bibliography base*) and can communicate them afterwards. The non-documented knowledge can be made of informal talks and ideas, hard-copies, further material or can even be recorded into obsolete digital media. P_i should make an effort in order to translate the non-documented knowledge into a digital format, via voice recording, notes, schemes, scanning hard-copies and so on.

The architecture in Fig. 3 evolves in time as shown in Fig. 4 and so does the documentation: the communication process itself and the (consequent or not) acquisition of further material restart the learning process and lead from state st1 to state st2, causing thus the evolution and enlargement of the partial bibliography base block.

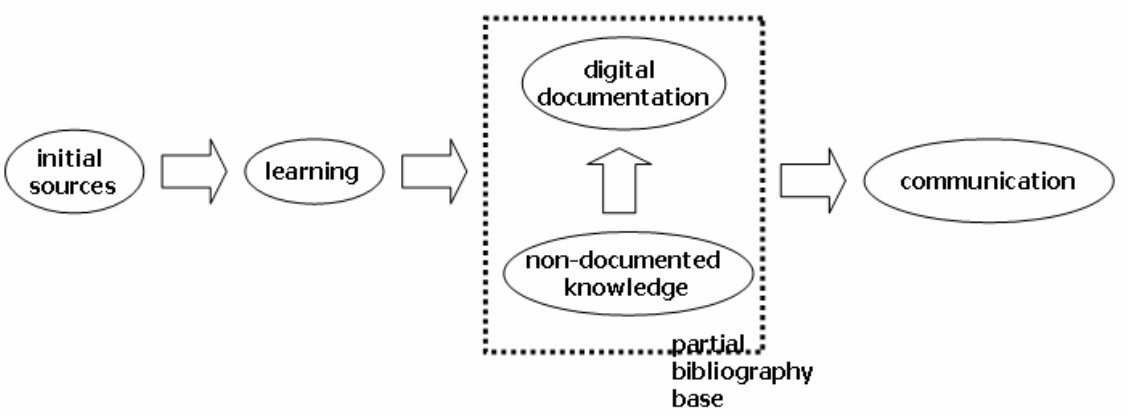

Fig. 3: from the initial sources to documentation and communication (*knowledge state* st_1 of P_i on S_j)

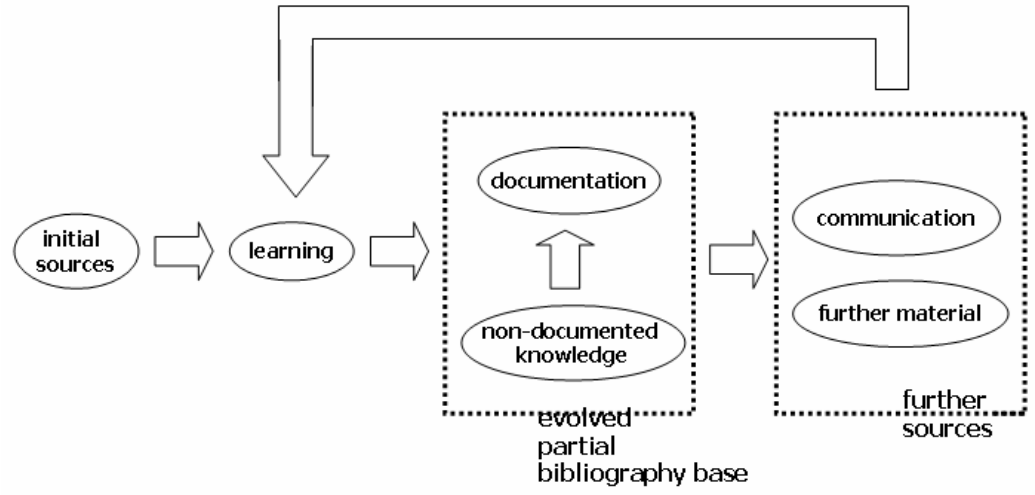

Fig. 4: knowledge evolution for P_i on S_j from st_1 to st_2, etc.

Let us now discuss each of these blocks in more detail. First of all, the learning block can be analysed as shown in Fig. 5, in order to classify the phases of knowledge evolution: the knowledge sources can be acquired and analysed by three main means: individual work, direct cooperation with people involved in the same subject or communication and knowledge sharing in general. This stage of knowledge processing can evolve in the study of related subjects or in a better understanding of the original sources, which can also interact and improve knowledge.

As far as the other blocks are concerned, the initial sources can be of many kinds: informal talks and ideas, notes, slides and papers from lectures, workshops or conferences, books, journals and so on, and should also contribute to the partial bibliography. The partial bibliography block, that will be discussed in more detail in the next subsection, can be roughly classified in more or less the same way and needs to be collected and structured on the basis of precise criteria, leading non-organised material to a structured layout (Fig. 6). The same applies to the object of communication and further material. The communication modalities are those proper of every kind of cooperation, ranging from papers, books, direct face to face talks, lectures and presentations, to real-time communication via traditional or broad-band services for cooperative work.

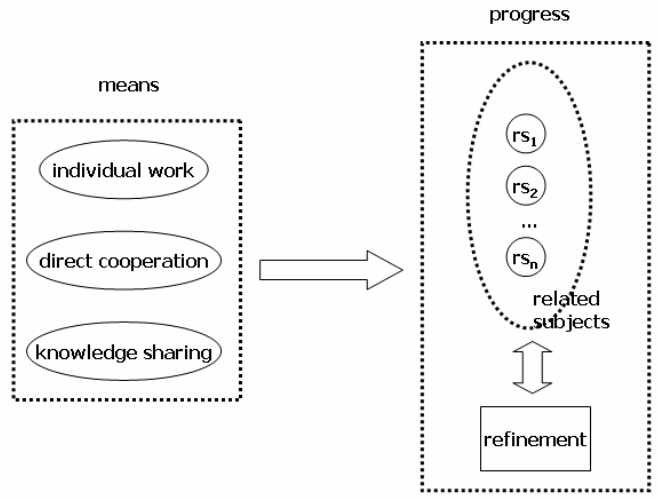

Fig. 5: learning evolution block

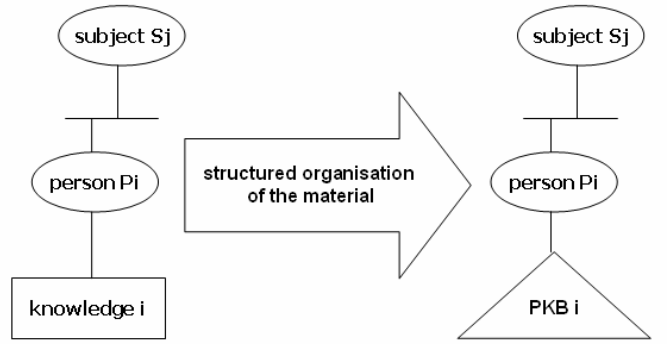

Fig. 6: organising material to build a partial bibliography

These considerations are the bases of the proposed architecture for the partial knowledge bibliography about S_j by a person P_i. In particular, the following features have been outlined:

[1] source documentation of many heterogeneous types is initially selected

[2] P_i produces documentation of many heterogeneous types and non-documented knowledge

[3] Pi organises his documentation using a structured digital model and layout, and tries to integrate his non-documented knowledge in such model

[4] During the learning process, P_i finds other people (authors, lecturers, colleagues, etc.) involved in the subject

[5] The process of learning evolves, improves knowledge, leads to the study of related subjects, enriches and refine the documentation

[6] The communication process also broadens, deepens and improves knowledge, and gives raise to links among people involved in S_j and make it necessary to revise the documentation

On the basis of these features, let us now investigate how the knowledge-bibliography can be defined. First, the structure of a partial-knowledge-bibliography and a possible implementation is defined. Then, the linking are depicted of many partial-knowledge-bibliographies made by different persons on the same subject, in order to build a knowledge-bibliography on that subject. Finally, the possibility is addresses of creating a higher-level clustered-layer containing many different subjects.

The Architecture for the Partial-Knowledge-Bibliography

The considerations in paragraph 2.2 will now be formalised in order to define a possible structure for the partial knowledge bibliography. We suggest LDAP (*Lightweight Directory Access Protocol*) [25, 26, 27] should be adopted, which provides both a modelling and an implementation tool, and we will refer directly to such structure. LDAP is used in an increasing number of applications, including enterprise databases, databases for storing network configuration information and service policy rules, storage of authentication rules and many others. LDAP is optimised for reading operations, so it could be very suitable for storing and managing bibliographies. Furthermore, the LDAP data model uses a hierarchy of classes, each class described by single-valued or multi-valued attributes. This tree structure allows to organise and navigate data in a very efficient, simple and user-friendly way. Furthermore, LDAP schemes can be very easily changed and extended in order to add new attributes and new classes. These operations, on the contrary, would be very time-consuming if traditional database systems were used. This feature can be very useful when designing a knowledge bibliography: as a matter of fact, new objects and new attributes are very likely to be modified or added, due to new needs, merged experiences or improvements made by the people who are developing it. LDAP was also built

for distributed environments, so it suits the distributed location of documentation very well.

The knowledge that a person acquires on a specific subject can be organized in an LDAP tree as shown in Fig. 7: the 0-level class describes the *subject* in general; the 1-level classes represent respectively the *source documentation*, the *produced documentation*, the *non-documented material*, expert *people* found during the learning and communicating process and some *related subjects*. The legenda of classes and attributes is detailed in Tab. 1 and describes the only attributes whose semantics can be ambiguous. It must be noticed that the multi-valued attributes *contacts* and *links* keep track of the links among documentation, authors and experts, related subjects and so on.

Even if many other types of classifications have been made and are available on the Internet, the aim of this proposal is to define a possible common structure for partial bibliographies as a basis for a creating a knowledge-bibliography on the considered subject.

Let us now thus face the problem of defining a common knowledge-bibliography from many partial ones.

Building a Knowledge-Bibliography by Merging Partial Bibliographies

Let us consider two people P_l and Pm who have prepared their own partial knowledge bibliographies PKB_l and PKB_m on the subject S_i. It is assumed that and PKB_m share the layout discussed in paragraph 2.3. Even if the structure is the same, PKB_l and PKB_m are bound to differ from many points of view: for instance, they can share some documentation, people and related subjects. In this case, especially when descriptions and personal judgements are concerned, such as relevance, state of the art, important links and so on, P_l and P_m must reach a common agreement or, at least, not add too many points of view. As far as the sources (such as links to available material, related subjects and contacts) are concerned, it must be decided to leave them all or select some of them. Other semantic problems, such as homonyms, synonyms, different styles should also be solved. On the basis of these considerations, the basic architecture for merging partial bibliographies is depicted in Fig. 8, with respect to PKB_l and PKB_m. The general case of merging many partial bibliographies is far more difficult: if many people do it simultaneously, the schema in Fig. 8 can be used. If further partial bibliographies are added later, the process is asynchronous and some coordinator should guide the integration.

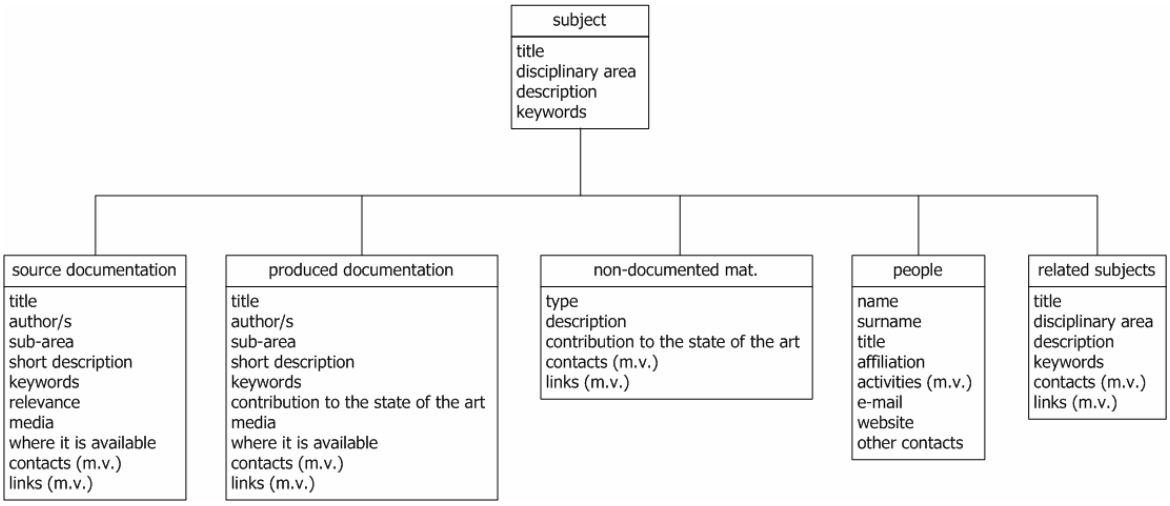

Fig. 7: an LDAP structure for the partial knowledge bibliography

class	attributes	description
subject	title, disciplinary area, description, keywords	
source documentation	title, author/s, sub-area short description, keywords, relevance media where it is available contacts (multivalued) links (multivalued)	 specific area within the subject personal opinion of the person on the material type of document and store: video-conference stored on a dvd, etc. url, private archive, etc. people related or cited suggested and other related links
produced documentation	title, author/s, sub-area	as above

	short description, keywords, contribution to the state of the art media, where it is available, contacts (multivalued), links (multivalued)	personal opinion of the person on his own contribution as above
non-documented-material	type description, contribution to the state of the art, contacts (multivalued) , links (multivalued)	note, ideas, informal talks, drawings, etc. as above
people	name, surname, title, affiliation, activities (multivalued), e-mail, websites, other contacts	specific interests in the subject
related subjects	title, disciplinary area, description keywords contacts (multivalued) links (multivalued)	people related or cited sites referring to related subjects

Tab. 1: classes and attributes of the partial knowledge bibliography

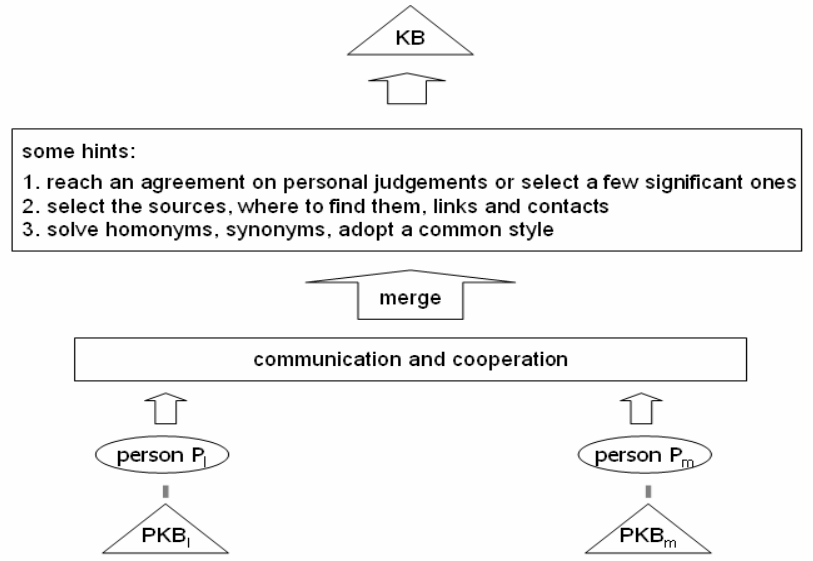

Fig. 8: merging two partial knowledge bibliographies

Clustering Knowledge-Bibliographies

A straightforward way to represent (not to make) many subjects and their own knowledge bibliographies is to come back to the layered graph model. Let us define a node as the couple (S_i, KB_i), where S_i is a subject and KB_i is its knowledge bibliography. Each edge is labelled with the description of the correlation.

The subjects can be clustered on the basis of their width: topics, sub-areas, area. For instance, in the database field, "relational databases" is a subject, "temporal databases" is a sub-area and "temporal SQL" is a topic. The symbolic layer contains topics, which can be clustered together if they belong to the same sub-area, the 1-clustered layer contains sub-areas, which can be clustered together if they belong to the subject and the 2-clustered layer contains subjects. Each item, again, can be linked to many different clusters and further levels of detail can also be defined.

3. CONCLUSIONS

In this paper some early architectures and methods were presented for the definition of a uniform bibliography on a subject. An LDAP-architecture was suggested for the definition of a *partial bibliography* made by a single person. Some early guidelines were stated for merging partial bibliographies into a common one. Further work will be devoted to the definition of protocols for the precise merging of documentation, in order to define a method for building a "global" and easily-understandable knowledge-bibliography. This structure must be shared by anyone via the Internet and also refined by adding new documentation on the basis of specific layouts.

4. REFERENCES

[1] Gratton L., Ghoshal S., "Improving the quality of conversations", **Organizational Dynamics**, 2002.

[2] Mengis J., Eppler M., "Understanding and Managing Knowledge-intensive Conversations", **#1/2005 ICA Working Paper, University of Lugano**, (USI), Switzerland, 2005.

[3] Eppler M, "The Concept of Knowledge Communication and Its Relevance to Management", **USI Research Note, Faculty of Communication Sciences, University of Lugano**, (USI), Switzerland, 2005.

[4] http://www.knowledge-communication.org/overview.html

[5] http://www.w3.org/Collaboration/Overview.html

[6] Eppler M., "Conceptual Management Tools: A Guide to Essential Models for Knowledge Workers", **USI Research Note, Faculty of Communication Sciences, University of Lugano** (USI), Switzerland, 2000.

[7] Woolf B. P., Reid J., Stillings N., Bruno M., Murray D., Rees P., Peterfreund A., Rath K., "A General Platform for Inquiry Learning", **Proceedings of the 6th International Conference on Intelligent Tutoring Systems**, Lecture Notes In Computer Science, 2002.

[8] Eppler M., "Knowledge Visualisation: Towards a new Discipline and its Fields of applications", **#2/2004 ICA Working Paper, University of Lugano**, (USI), Switzerland, 2004.

[9] van Harmelen F., "Practical Knowledge Representation for the Web", **AI-ED 2001 Workshop**, 2001

[10] Schamber L., Bateman J., "User Criteria in Relevance Evaluation: Toward the Development of a Measurement Scale", **ASIS&T Annual Conference**, 1996

[11] Barry C. L., "User-defined relevance criteria: an exploratory study", **Journal of the American Society for Information Science, ACM**, 1994

[12] Tombros A., Ruthven I., Jose J. M., "How users assess web pages for information seeking", **Journal of the American Society for Information Science, ACM**, 2005

[13] Science Citation Index, Istitute for Scientific Information

[14] Pacuit E., Parikh R., "The Logic of Communication Graphs", **Proceedings of. Declarative Agent Languages and Technologies (DALT 2004), Lecture Notes in Computer Science**, 2004

[15] Hosobe H., "A high-dimensional approach to interactive graph visualization", **SAC'04**, 2004

[16] Zhang L., "Knowledge Graph Theory and Structural Parsing", **Ph.D. thesis, University of Twente (UT)**, 2002

[17] Trzaska M., Subieta K, "Structural Knowledge Graph Navigator for the ICONS prototype", **IASTED International Conference on Databases Applications**, 2004

[18] http://www.hcrc.ed.ac.uk/ilex/

[19] Maio D., Rizzi, S. , "Layered knowledge architecture for navigation-oriented environment representation". **TR CIOC-C.N.R., n. 108**, 1996

[20] C. De Castro, "A Temporal Layered Knowledge Architecture for an Evolving Structured Environment", **Proc. of 27th Int. Conference SOFSEM**, Milovy (Brno), Czech Republic, 2000

[21] Spector L., Hendler J., "Knowledge strata: Reactive planning with a multi-level architecture", **TR UMIACS-TR-90-140, UM Institute of Advanced Computer Studies, Univ. of Maryland**, 1990

[22] Signore O., "The Layered Architecture of Semantic Web", http://www.w3c.it/talks/eva2004Jerusalem/

[23] Tim Berners-Lee, "Semantic Web Roadmap", http://www.w3.org/DesignIssues/Semantic.html

[24] Eades P., Lin Huang M., "Navigating Clustered Graphs using Force-Directed Methods", **Journal of Graph Algorithms and Applications**, 2000

[25] W. Yeong, T. Howes, S. Kille, "Lightweight Directory Access Protocol" **IETF RFC 1777**, http://www.ietf.org/rfc/rfc1777.txt, 1995.

[26] Howes T., Smith M., Good G., Understanding and Deploying LDAP Directory Services, **Addison Wesley** 2003, 2^ ed.

[27] D. Kandlur, H. Schulzrine, D. Verma, X. Wang, "Measurement and Analysis of LDAP Performance", **Network Systems Department - IBM** T.J. Watson Research Center, 1998